PATENT ABSTRACTS OF JAPAN

(11)Publication number:

09-146842

(43)Date of publication of application: 06.06.1997

(51)Int.CI.

G06F 12/08

G06F 12/08

G06F 3/06

G06F 3/06

(21)Application number: 07-300967

(71)Applicant: HITACHI LTD

(22)Date of filing:

20.11.1995

(72)Inventor: KOBAYASHI RIE

MATSUMOTO YOSHIKO

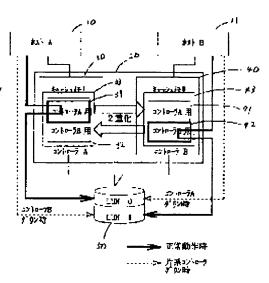
MURAOKA KENJI

(54) STORAGE SUBSYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide the storage subsystem with a controller which eliminates exclusive control over a cache between plural controllers sharing the cache.

SOLUTION: The cache areas of caches 33 and 43 in which data are written mutually in multiple are divided by processors and the controllers 30 and 40 access only their controller control areas. The cache areas that the controllers use are fixed to eliminate the need for exclusive control between the processors and prevent deterioration in performance due to multiprocessor constitution.



LEGAL STATUS

[Date of request for examination]

31.10.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

配信サブシステム

(19)日本四分表庁() P)

(12)公開特許公報(A)

(11)特許出記公司等号 特開平9-146842

輪間平9-146842

(43)公額日 平成9年(1997)6月6日

(51) Int.CL*	機別配号	广内装理器号	Fl		技権表示舞所
C06F 12/06	3 2 0	7623 - 5 B	G 0 6 F 12/08		320
		7623 - 5 B			į
		7623 - 5 H			н
3/06	302		3/06 3 0 2 A		
	304		304C		304C
			警查請求	未請求	網求項の数19 OL (全 14 頁)
(21)出業書号	特數平7 - 300967		(71)出版人 000005108		
			t	集式会	社目立動作所
(22) HSMR B	半成7年(1995)11月20日		1	東京都	千代田区神田駿河廿四丁目 6 香地
			(72) 免明者	小林	村坐
			ļ	神奈川	从小田原市国内体25610多地 株式会
			į	社日立	製作所ストレージシステム事業部内
			(72) 発明者	松本	住于
			1	神奈川	県小田原市国府坪2680番地 株式会
				社口立	製作所ストレージシステム事業部内
			(72) 発明者	村開	差 57
			1	神多川	体小切原市国内牌2880番地 株式会
			1	社门立	製作所ストレージシステム事業部内
			(74)代理人 弁理士 小川 製勇		

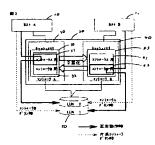
- 1 -

(54) 【発明の名典】 配管サブシステム

(57) (事約)

【課題】キャッシュを共有する複数のコントローラ間の キャシュの排色制御をなくした制御装置を有する記憶サ ノステムを提供する.

/// ひを提供する。。 【解決手段】互いに多意書きされているキャッシュ3 3、43において、キャッシュ領域を各プロセッサ毎に 分割し、各々のコントローラ30、40は、自コントロ - ラ餅御エリアのみにアクセスする。各コントローラが 使用するキャッシュ領域を固定化することにより、プロ セッサ間の禁機制御を不要とし、複数プロセッサ化に伴 う性能劣化を防止することが可能となる。



2003 04 24 10.53

(特許請求の範囲)

【競送取1】 ホストコンピュータのデータを終終し、数 数の記憶的域を有する記憶装置と、

故ホストコンピュータの指示に基づいて数記憶装置の納 御を行い、数ホストコンピュータと数ディスク装置との 間のデータ伝送を制御し、数ホストコンピュータと数配 像装置との間を転送されるデータを一時的に保持する役 数の値域を有するキャッシュメモリを有する役数のコン トローラと前記複数のコントローラ間を接続するパスと を具備する制御装置とそ有する記録サブシステムであっ

前記コントローラには、前記記憶装置の複数の記憶領域 のうち少なくとも1つと数コントローラのキャッシュメ モリの複数の領域のうち少なくとも1つと解記パスによ り接続される他のコントローラのキャッシュメモリの収 世の領域のうち少なくとも一つが割当てられることを特 数とする記憶サブンステム。

【請求項2】 請求項1記載の記載サプシステムにおい て、前記コントローラは、前記ホストコンピュータから 転送されるデータを放コントローラに割当てられている。 複数の前記キャッシュメモリに書込むことを特徴とする 起懐サブシステム。

【翻求項3】翻求項2記載の記憶サブシステムにおいて、前記コントローラに障害が発生したときは、前記也 のコントローラは鉄整賞コントローラが担当していた値 紀紀娘装置の紀像領域の処理を行うことを特徴とする紀 使サブシステム。

「独求項4) 独求項3.紀保の記憶サブシステムにおい て、鉄也のコントローラはホットスタンパイしているコ ントローラであって、ホットスタンパイしているコント **ーラには、キャッシュメモリの記憶領域を割り当てな** いことを特徴とする記憶サブシステム。

【請求項5】請求項1記数の記憶サブシステムにおいて、前記制御装置は複数の前記コントローラ間を接続す **あパスをむし、必知コントローラが他のコントローラに 割当てられた前記記像装置の記憶領域に対する処理要求** をホストコンピュータから受取ったときは、前記コント ローラは、前記他のコントローラに放処理要求を通信することを特徴とする記憶サブシステム。

【請求項6】 請求項1記載の記憶サブシステムにおい て、前記キャッシュ領域の分割は、コントローラの負荷 に応じて変更することを特徴とする記憶サブシステム。 【翻求項7】 ホストコンピュータのデータを格納する複数の論理ポリュームを有する磁気ディスクと、

終ホストコンピュータと数ディスク装置との間を転送さ れるデータを一時的に保持する複数の領域を有するキャ ッシュメモリと、前記キャッシュメモリとが接続され、 技データのデータ転送を制御するデータ転送制御服とを 有する複数のコントローラと、複数のコントローラ間を 基づいて鉄磁気ディスク装置の制御を行う制御装置とを 存する記憶サブシステムであって

前記コントローラには、前記磁気ディスク装置の複数の 高理ポリュームのうち少なくとも1つと数コントローラ

のキャッシュメモリの複数の領域のうち少なくとも1つ と、他のコントローラのキャッシュメモリの複数の領域 のうち少なくとも!つとが割当てられることを特徴とす

○記録リフンステム。 【請求項8】 請求項7記載の記憶サブシステムにおい

て、作記コントローラは、俗記ホストコンピュータから **以送されるデータを数コントローラに割当てられている** 終コントローラのキャッシュメモリの領域と、終コント ローラに割当てられている他のコントローラのキャッシ ュメモリの仰城とに書込むことを特徴とする記憶サブシ

て、仮記コントローラに練客が発生したときは、府記包 のコントローラは鉄路省コントローラが担当していた前 記論時ポリュームの処理を行うことを特徴とする記憶サ プシステム.

【雌求項10】 額求項9記載の記憶サブシステムにおい て、貞妃他のコントローラはホットスタンパイしている コントローラであって、、ホットスタンパイしているコ ントローラには、キャッシュメモリの記憶領域を割り当 てないことを特徴とする記憶サブシステム。

【請求項11】 請求項7記載の記憶サブシステムにおい て、前記コントローラが他のコントローラに割当てられ た会理ポリュームに対する処理要求モホストコンピュー タから受けとったときは、前紀コントローラのデータ転 決制機能は、放他のコントローラに自記第一のバスを介 して処理要求を転送し、放処理要求を受領した該他のコ ントローラが抜編理ポリュームに対する処理を行い、処 理結果を、前記コントローラに転送することを特徴する

【韓菜項】21 韓菜項7 紀載の記載サブシステムにおい て、前記キャッシュ領域の分割は、コントローラの負荷 に応じて変更することを特徴とする記憶サブシステム。 【趙求項13】趙求項7記載の記憶サブシステムにおいて、前記コントローラ間のパスは、2つの前記コントロ

ーラを接続する第一のパスと、、数 2 つのコントローラ の組を接続する第二のパスを含むことを特徴とする記憶 サブシステム。

「趙宝斯) 4) 趙宝斯(3記載の記憶サブシステムにお いて、前記制御装置にコントローラを増設するときは 前記コントローラの2台単位に用設することを特徴とす

る記憶サブシステム。 【舘求項15】ホストコンピュータのデータを格納し、 複数の記憶領域を有する記憶装置と、

数ホストコンピュータの指示に基づいて数記憶装置の制 接続するパスとを有し、彼ホストコンピュータの指示に 50 御を行い、故ホストコンピュータと故記憶装置との間の

2003 04 24 10.53

記憶サブシステム

特開平9-146842

データ転送を制御し、数ホストコンピュータと放記数姿 間との間を転送されるデータを一時的に保持する複数の 低減を存するキャッシュメモリを有する複数のコントロ 備する制御装置とを有する記憶サブシステムであって、 前記コントローラには、前記記憶装置の複数の記憶領域 のうち少なくとも1つと終コントローラのキャッシュメ そりの複数の領域のうち少なくとも1つと前記パスによ り接続される他のコントローラのキャッシュメモリの複 酸の個域のうち少なくとも一つが割当てられ、

前記ホストコンピュータから転送されるデータは、終コ ントローラに割当てられている終コントローラのキャッ シュメモリの価値と、終コントローラに割当てられてい る他のコントローラのキャッシュメモリの領域に書込まれることを特徴とする記憶サブシステム。

【翻求項16】翻求項15記載の記憶サブシステムにおいて、前記コントローラに降省が発生したときは、前記 他のコントローラは故障害コントローラが担当していた 前記記憶装備の記憶領域の処理を行うことを特徴とする 記憶サブシステム。

【離求塔】7】請求項15記載の記憶サブシステムにお いて、鉄色のコントローラはホットスタンパイしている コントローラであって、ホットスタンパイしているコン トローラには、キャッシュメモリの記憶領域を割り当て ないことを幹徴とする記憶サブシステム。

「神像原181 神像原15記録の記憶サブシステんにお いて、自記制御答置は複数の前記コントローラ間を接続 するパスを有し、餡配コントローラが色のコントローラ に割当てられた前記記憶装置の記憶領域に対する処理製 求をホストコンピュータから受取ったときは、前記コン トローラは、直紀前のコントローラにお処理を求を通信 することを特徴とする記憶サブシステム。

【静波項19】 離球項15紀歳の紀像サブシステムにお いて、前記キャッシュ領域の分割は、コントローラの負 荷に応じて変更することを特徴とする記載サブシステ

【発明の属する技術分野】 本発明は、上位装置からの情 報の入出力要求を制御する制御装置を有する記憶サブシ ステムに関し、特に、製御装置内のコントローラ及びも ヤッシュメモリを冗長構成とする記憶サブシステムに関

[0002]

【従来の技術】コントローラ及びディスク多の記憶装置 に冗長性を持たせた記憶サブシステムとして、 一へ、ICE TACULAC 第7 ノンステムとして、 一方の状が型用系として、他方の系が干燥系として投資する2 意の系で構成される記憶サブシステムがある。

【0003】特徴平4-215142に記載されている 記憶サプシステムは、現用系のディスク装置の記憶情報 50 Aでは、この排始制御により、シングルコントローラの

を由某からアクセス可能な共用ディスク装置を介して予 備系のディスク装置に複写すること、あるいは、規用系 コントローラ酸省時は、予備系のコントローラによっ て、現用系のディスク装置の記憶情報を抽出可能とする ことによって、コントローラ及びディスク装置障害時の 05

データ保全性の向上を計っている。

(発明が解決しようとする課題) 最近の市場動向とし

て、記憶装置の高性能化、大容量化、低価格化の要求が 高まっており、RAIDの技術が重複されている。RA I Dの技術を適用した記憶サブシステムにおいては、複 数のディスク装置をアレイ状に構成する。そして、データ書き込み時には、書き込みデータに加えて冗長データ ・ を書き込みデータを格納したディスク装置とは異なるデ

ィスク複数へ響き込む。アレイ構成内の任意のディスク 装置の故障に対しは、他のディスク装置のデータと前記 冗長データから障害ディスク装置上のデータを修復可能 とすることによって、ディスク装置のデータの保全性の 向上を計っている。

【0005】しかし、RAIDの技術を適用した記憶サ プシステムは、データの保全性が向上する反面、上述し た冗長データ生成。「書き込みのために処理時間が増大す るため、ホストからの1 ()処理と同期して冗技データ の生成、書き込みまでを行うと、ホストからのライト作 能が大幅に劣化する、従って、RAIDの技術を適用し **Z**5

た記憶サプシステムのコントローラには、ライトキャッ シュが不可欠となっている。 【0006】ライトキャッシュとは、コントローラ内に

構成された。データを一時的に書き込むキャッシュであ り、ホストからのライトを求では、このキャッシュに妻 き込みを行った時点で、ホストに終了報告を返す。そし て、ホストの1 O処理とは非同期に冗長データの生成、ライトデータ及び冗長データのディスク装置への格 ぬを行うことにより進き込み供養的の性能低下を妨ぐ。

しかし、ライトキャッシュを用いると、データをキャッシュ上におき込んだ時点でホストに終了報告をするた の、キャッシュ上にディスク装置大反映のホストデータ が存在する。従って、キャッシュに几長性がなければ、 キャッシュ練書時に、ユーザデータロストとなる。した がって、特にデータの高は似性が強く求められる記憶サ ブシステムに用いる制御装置では、従来のコントロー ラ、記憶装置の冗長構成に加え、一般的にキャッシュに も冗長性を持たせることが行われている。

【0007】コントローラを多数化した記憶サブシステ ムにおいて、単にキャッシュを多重化すると、キャッシュ上のデータを複数の制御装置から同時にアクセスする ことによるデータ整合性矛盾を防ぐためにキャッシュア クセス時に複数の新額装置からの基地制御が必要になる。そして、コントローラを多様化した記憶サブシステ

記憶サブシステム

特開平9-146842

記憶サブシステムに比べて性能が低下する。

【0008】 本発明の目的は、コントローラの多重化及びキャッシュの多重化に伴う、プロセッサ間のキャッシ ュの排他制御を無くし、性能を痛とすことなく信仰性を とげることにある。

(00009)

・ (舞瞬を解決するための手段)上紀の目的を達成するた め、本発明による記憶サブシステムは、各プロセッサ毎 に飢買相当の論理ポリュームを排他的に決める手段と、 あるプロセッサが受領したホストコンピュータからの要求が、担当外であった場合は、担当プロセッサに処理要 求を適信する手段と、上記通信を受領したプロセッサ は、処理結果を要求元プロセッサに通信する手段と、 プロセッサ低に、ディレクトリ、データセグメント等の キャッシュ構成要素を持つ手段と、上記構成要素の状盤 をプロセッサの負荷に応じてダイナミックに変更する手 段と、ホストコンピュータからのライトデータを複数の コントローラ上のキャッシュへ多重書きする手段と、コ ントローラ報告時には、解告コントローラ内プロセッカ の持つキャッシュ構成要素の制御権を正常系コントロー ラ内のプロセッサに切り替える手段と、コントローラ復 田時には、政制調査を提出プロセッサに成す手段と、デ ィスク装置への書き込み時にキャッシュメモリ酸害が発 生した際は、多重要をしている他キャッシュからディス ク装置に書き込みを行う手段とを有する。 【0010】上述した手段により、複数のプロセッサ間

で、キャッシュを排他制御することなく複数のコントロ --ラ上のキャッシュへ多重書きを行うことができ、複数 プロセッサ化に伴う性能低下の発生を防ぎ、性能を落と

テントにより、100mmの トラスト マンス・スター マンス くら 新性の向上を計ることができる。 【0011】 また、上述の手段により、キャッシュメモリ酸省時には、多重書きしている他キャッシュからディ スク装置への書き込みを行い、データロストを防止でき

【0012】さらに、七紀手段により、コントローラ牌 害時には、自動的に、正常系に切り養えて処理破行が可 能であり、また、コントローラ復和特には、自動的に · 福田系に観撃を戻すことが可能となり、システムの無停 止運用を実取できる。

(0013) · 【発明の実施の影盤】例上は、本発明の最急間である

【0014】図1において、10、11はホストコンピュータ、20はデュアルコントローラ構成をとる対策値 数、50はディスク装置であり、ディスク装置50は、 金埋ボリューム0と高度ボリューム1の2つの盒厚ボリ 一ムに分割されている。

【0 0 1 5】ホストAIOは、初間答案での内のコント ローラム30を介して、為程ポリューム0の色質を行っ ており、ホストB11は、納朗協器20内のコントロー 10日間のデータ転送を制御す ラB40を介して、森代ボリューム1の処理を行ってい 50 0、6000より構成される。

【0.0.1.6】ここで、コントローラA.3.0には論理ポリ ューム Oが、コントローラ B 4 D には島境ポリューム i が処理担当島度ポリュームとして割当てられている。

又、コントローラ内のキャッシュの領域は、それぞれ、 コントローラA用キャッシュ31、41、コントローラ B用キャッシュ32、42に2分割されている。そして、コントローラA用キャッシュ31と41の間で2重 書きを行い、又、コントローラB用キャッシュ32と4 2の間でも2重要さを行う。 【0017】コントローラA30は、通常、コントロー

ラA用キャッシュ31と41を用いて、1, O処理を行 い、同様に、コントローラB40は、コントローラB用 キャッシュ32と42を用いて、1,/O処理を行う。こ のように、コントローラ毎に使用するキャッシュ 領域を 観別に割り当てることにより、コントローラ目の排他制 脚を無くし、コントローラ台数増加に伴う性能的化を妨

【0018】また、コントローラお40粒害時には、コ ントローラB用キャッシュ32、42をコントローラA 30が使用することにより、ホストA10からコントロ --ラA30を介して、コントローラB40の処理科当で

あった倫理ポリューム1への処理を統行させることがで 【0019】以下、本発明によるマルチコントローラ機 岐の制御装置のし実施例を図面を用いて説明する。

【0020】図2は、本発明をマルチコントローラ構成 の磁気ディスクアレイサブシステムに適用した場合の構 成因である. [0021] M2EHUT. 1000. 1100. 12

00、1300はデータ処理を行う中央処理体装置である ホストコンピュータ、2000はマルチコントローラ網 成モとりディスク装置の制御を行う制御装置、700 ペニュンフィスン学品の物質を行う制度を置って00 0、7100はホストコンピュータのデータを格的する ディスク経営である。ここで、物質接営2000は、ホ ストパスに直結したスロットに及し込みホスト程体内に

組み込む場合もあるし、制御装置として独立した物体に 組み込む場合もあるし、ディスク装置を組み込んだ僚体 として実現する場合もある。また、ディスク装置料70 00及び7100は、データディスクとパリティディス

りからなるパリティグループを含んでいる。 さらに、テ ィスク装置群7000は、論理ポリューム0と論理ポリ ューム!とに、ディスク装置群7100は麻煙ポリュー ム2と為数ポリューム3とに分割されている。 【0022】 初脚装置2000は、ホストコンピュータ

1000、1100とディスタ協議7000間のデータ 転送を制御するコントローラ3000、4000及びホ ストコンピュータ1200、1300とディスク協議7 1 6 0 間のデータ伝送を製御するコントローラ5 0 0

【0023】コントローラ3000は、ホストコンピュ ロセッサ (以下 (プロセッサ) という。) 3200、デ ータの転送を実行するデータ転送制御服3300、ホストコンピュータ1000とディスク装置7000のデー タ転送時及びプロセッサ間通信時に用いられるキャッシ ュ3400、各ディスク装置7000とのプロトコル制 器を行うDRVI、下制御部3500より構成される。

コントローラ4000、5000、6000はコントローラ3000と同一の構成である。 【0024】プロセッサ3200は、後述のF12によ り、あらかじめプロセッサ毎に禁錮的に努り当てた刑当 無理ポリュームの処理を行う。このプロセッサ毎の利当 無理ポリュームの指定は、ホストコンピュータから適理 ポリューム毎の担当プロセッサ指定コマンドを受け取る ことにより、ダイナミックに収定可能である。このプロ セッサと担当論理ポリュームとの対応情報は、後述のキ ャッシュ上の共通メモリ蝦蛄3410.4410に移納

【0025】データ転送制御部3300はプロセッサ3 200からの指示により、ホストコンピュータ1000からのライトデータを指定キャッシュに多重書きする機 節を備えている。この実施例の構成では、キャッシュ: 400とキャッシュ4400の間で2歳書きを行い、 た、キャッシュ5400とキャッシュ6400の間でも 2章書きを行う、以上、キャッシュ3400とキャッジ 14400の2面に2重要をする方式について説明す

100261 ****************** の内容について図3を用いて説明する。ぬキャッシュ3 400とキャッシュ4400は内障機成が同一であるた キャッシュ3400を例に説明する。キャッシ 400は、プロセッサ関連値に用いる紡績情報を格納し ている共通メモリ領域3410、プロセッサ3200月 飯域3480、プロセッサ4200用飯域3490より 構成される。

【0027】プロセッサ3200用領域3480は、オ ストコンピュータとディスク装置間のデータ転送時、データを1次的に格納するデータ格納エリア3482、デ ヤッシュ4400内のプロセッサ3200用領域448 0に2度費きを行う。同様に、プロセッサ4200用領域3490は、プロセッサ4200により、キャッシュ 4 4 D D内のプロセッサ 4 2 D D 用級域 4 4 9 Dのライトデータとライトデータの管理情報が2 重着きされてい

ム担当プロセッサ機製3420、プロセッサ負債機製3 430、多重器を情報3450、プロセッサ間コミュニ ケーションメモリ3460より構成され、これらの情報 は全て、データ転送網舗舗3300、4300によっ て、キャッシュ3400と1400に2項目をされてい

特闘平9-146842

【0029】 図3 (c) にプロセッサ間コミュニケーシ ョンメモリの構成を示す。プロセッサ間コミュニケーションメモリの構成を示す。プロセッサ間コミュニケーションメモリ3460は、プロセッサ3200、420

0、5200、6200 6時の過ぎ込み用メモリ34セ 1、3462、3463、3464より構成される。 3(d)にプロセッサ響き込み用メモリの構成を示す。 プロセッサ3200番を込み用メモリ3461は、日ブロセッサ以外のプロセッサ4200、5200、620

8への要求用エリア3471,3472,3473と行 プロセッサ以外のプロセッサ4200,5200,62 00からの要求に対する必番用エリア3474,347 5、3476より構成される。プロセッサ4200、5 200、6200mを込み用メモリ3462、346 3、3464の内部構成は、プロセッサ3200mを込

み用メモリ3461と同一構成である。 【0030】キャッシュ5400とキャッシュ6400 との関も、共通メモリ領域を除いて、キャッシュ340

のとキャッシュ4400との間と阿様に2重化が行われている。共通メモリ蜘蛛は、キャッシュ3400、44 **0.0に2乗者きされている情報を制御装置内の全プロセ** サで共有するため、キャッシュS400、6400に はかをしない.

【0031】 本発明を実施する制御装置では、コントロ マラの物表はコントローラ2台単位で行い、早になった コントローラのキャッシュ間のみで2番音を行うとと もに、ドライブ値のデータバスについても、それぞれの ディスク装置は対になったコントローラにのみ接続する ことによりハードウェア構成を商略化し、ドライブ側デ ータバストの競合を回避することが可能となる。

【0032】次に本実施例における、磁気ディスクサブ システムでの、ホストコンピュータ1000からの1. O処理について関4、例5、例6を用いて控制する。± プロセッサ3200担当路理ポリュームへの L/O処理について説明する。

の書き込み要求時、プロセッサ3200は、まず、共通 メモリ領域3410内の論理ポリューム担当プロセッサ 情報3420によって、処理要求論理ポリュームの担当 プロセッサ情報を取得し、自処理担当論理ポリューム (LUN) への処理かの判定を行い(ステップ90

2)、自プロセッサ処理担当論様ポリュームへの処理で る。 あることを建成する。次に、処理権別の利定を行い(ス 【0028】共通メモリ領域3410は、論理ポリュー 50 デップ903)、巻き込み処理であることを認識する。

2003 04 24 10 53

ホスト1、F朝御部3100により、誰さ込み益物デー タを受難し、データ転送制御部3300によってキャッシュ3400のコントローラ3000用取破3480と キャッシュ4400のコントローラ3000用質減44 8 0 とにその管理情報とともに2重に格納する(ステッ ブ904)、そして、この時点でホストロンピュータ1 000に終了を報告する(ステップ905)。 【0034】図5は、キャッシュ内のデータをディスク 装御に格納する処理を示すフローチャートである。ブロ

セッサ3200は、ホストコンピュータ1000からの *O処理とは非同期にプロセッサ3200用軸編34 50上のライトデータをデータ転送制御配3300とD RV 1、下制提展3500によりディスク装置数70 00に格納する(ステップ922)。この数、キャッシ 3.のメモリ健実により並み込みエラーが発生した場合 (ステップ923) は、2重化しているプロセッサ32 00川銀城4480からディスク装置7000へ格納す る (ステップ924) ことによりデータ損失を妨止する ことができる。

【0035】ポストコンピュータ1000からの設み込 み要求時は、プロセッサ3200は、上記書き込み処理 阿様、自プロセッサ処理相当論理ポリューム(LUN) への処理であることを認識(ステップ902)した権 処理種別の料定を行う(ステップ903)。 1. 〇処理 が鉄み込み処理であることを記載すると、データ転送制 構成3300とDRV 1.1F動機成3500によりデ タモディスク装置群700Dからキャッシュ34 のコントローラ3000月酸暖3480に格納し(ステ ップ906)、ホストコンピュータに転送する (ステッ

/ 9 0 7 7 。 【0036】 次にホストコンピュータ 1 0 0 0 からコン トローラ4000個当島理ポリュームへの E. O処理に ついて裁明する。

【0037】ホストコンピュータi000からの書き込 み要求時、プロセッサ3200は、まず、共通メモリ勉 號3410内の論理ポリューム担当プロセッサ情報34 20によって、処理要求論理ポリュームの担当プロセッ 20によって、処理数求額様でリュームの担当プロセック特徴を収録し、自動提供加速が増加し、の利理かの料定を行い(ステップリロ2)、処理担当外高理ポリュームの影響であることを認識する。次に、影響を入り、対策であることを認識する。そして、ホストコンピュータ1000 からの書き込み塾理データをキャッシュメモリのコン) ーラ3000用領域3480に格納し、書き込み処理 をこの論理ポリュームの担当であるコントローラ400 0へ要求する (ステップ 9 0 9) 。 (0 0 3 8) プロセッサ3 2 0 0 は、プロセッサ4 2 0

0に着き込み処理を要求するために、、着き込みデータ 森度アドレス、着き込みデータのキャッシュ上の格納フ ドレス、データ長及び処理権別情報をデータ転送制御郎 50 処理担当であるプロセッサ4200に放み込み要求を過

3 3 0 0 により共通メモリ領域 3 4 1 0 、4 4 1 0 内の プロセッサ3200番を込み用メモリ内のプロセッサ4 200への要求用エリアに2項に格納する。ここで、他 理権別情報とは、最き込み処理が扱み込み処理がを判断 する情報である。プロセッサ4200は、例えば10m s といった一定時間で、共通メモリ蜘蛛3.4.1.D.

1 0の自プロセッサへの要求用エリアを基別にいき、他 プロセッサからの要求を認識する。

【0039】関6は、プロセッサ3200からの処理を 求を受けしたときのプロセッサ4200の処理をポすフ ローチャートである。相述の方法により、プロセッサ3 200からの要求を提直 (ステップ931) したプロセ ッサイ200は、プロセッサ3200歳を込み用メモリ 内のプロセッサイ200への要求用エリア内の処理権別

を参照し、書き込み処理会末であることを認識する(ス テップ932)、そして、プロセッサ4200は、プロ セッサ3200番き込み削メモリ内のプロセッサ420 0 への長塚州エリア内の書き込み益理アドレス。書き込 みデータのキャッシュ上の格納アドレス。データ技を助 得し (ステップ933)、キャッシュ3400内の維格

納アドレスからデータ長分の書き込みデータをプロセッ サイ200用領域3490と4490に、その管理情報 である書き込み論理アドレスとデータ長と共に、2重に 格納する(ステップ934)。そして、終了情報を共適 メモリ領域3410、4410内のプロセッサ4200 **書き込み用メモリ内のプロセッサ3200からの要求に** 対する応答用エリアに設定することにより、プロセッサ 3200に処理終了を通信する(ステップリ35)。

【0040】プロセッサ3200は、プロセッサ420 むに対する処理要求後は、プロセッサ4200番を込み 州メモリ内のプロセッサ3200からの要求に対する店 各用エリアを参照することにより、プロセッサ4200 の処理の終了を監視 (ステップ910) しており (図4

の鬼味の終了を監視(ステップリーリ)しており、1584年 参照)、この処理終了の適信を受けて、ホストコンピュ 一夕1000に終了を報告する(ステップ905)、プ ロセッサ4200は、この後、図5に従ってホスト!。 〇処理とは非国間に、この書き込みデータのディスク協 000への書き込み処理を行う。

【0041】 倒4において、ホストコンピュータ100 0から読み込み要求があったときは、プロセッサ3200は、上配書き込み要求受損時向様、処理担当外籍理ポ リューム (LUN) への処理であることを認識した (ス テップ902) 後、処理種別の料定を行う (ステップ9 08) 、 読み込み処理であることを認識すると、プロセ ッサ3200は放み込み要求益程アドレス、放み込みデ 一夕のキャッシュ上の格納許可アドレス、データ技、処 理権別情報を共通メモジ領域3410、4410内のフ

配催サブシステム

特別平9-146842

信する (ステップ911)。

【0042】図6において、プロセッサ3200からの 要求を認識(ステップ931)したプロセッサ4200 は、共通メモリ個域内の情報により、読み込み処理であ ることを認識する(ステップ932)。そして、共通メ そり領域から読み込み要求論理アドレス、読み込みデー タのキャッシュ上の格納許 リフドレス、データ技を収荷 する(ステップ936)、次に、データをディスク装置 7000からプロセッサイ200附領域イイ90に格納 し、このデータをキャッシュ3400上の格別作可アド レスに格割する(ステップ937)。さらに、共通メモ リ領域3410、4410内のプロセッサ4200番を 込み用メモリ内のプロセッサ3200からの要求に対す る応答用エリアに終了情報を設定することにより、プロ セッサ3200に鉄み込み終了を通信する(ステップ9

【0043】例4において、プロセッサ4200の処理 終了を監視 (ステップ912) していたプロセッサ32 00は、この親み込み終了報告を受けて、データをホストコンピュータに転送する(ステップ9 | 3)。

【0044】 このように、プロセッサ3200は、通 常、プロセッサ3200用領域3480と4480を用いて、1. (O処理を行う、同様に、プロセッサ4200 は、通常、プロセッサ4200用の終3490と449 0を用いて、1./O処理を行う。 {0045}このように、プロセッサ毎に使用するキャ

ッシュ領域を固定化することにより、プロセッサ間の鉄 数制御を無くし、プロセッサ台数増加に伴う性能劣化を 紡ぐことができる。特にホストコンピュータ間でファイ ル (急度ポリューム) モシェアしないシステムにおいて は、接続しているコントローラ内のプロセッサにこの会 理ポリュームを割り当てておくことにより、 1.10 処理 のときのプロセッサ間の適併制御を不登とし、さらなる 性能向上を可能とする。

【9046】次にコントローラ4000の間割時の自動 切り巻大、復刊方式について図7、図8を用いて成明する。1、〇処理実行中、コントローラ4000の超宵を 検知したプロセッサは、共通メモリ飢壊3410を用い C. 残りの全プロセッサにコントローラ4000の原告 を適信する。この際、コントローラ4000とキャッシ ュモ2重要をしているコントローラ3000内のブロヤ ッサ3200には、処理の引動要求も適保する。 4次編 例では、プロセッサ3200が陥害を検知した場合につ いて反明する。

(0047] 例7は、プロセッサ3240がコントロー 4000の障害を検知した場合のプロセッサ3200 の観覧を示すフロッチャットである。プロセッサ320 0は1、「〇見理案打中(ステッツ950)、コントロッ ラ4000の解答を使知(ステッツ951)すると、森 迷の方法により、プロセッサうでもり、6200にコン

トローラ4000の障害を通信する。そして、障害コン トローラモシステムから切り放すため、キャッシュ34 00と4400へ2重奏きされているホストコンピュー タからの書き込みデータ及び共通メモリ領域のデータ を、キャッシュ3400への1番番さに変更することを データ転送制御練3300に指示する (ステップ95 2)。また、プロセッサ3200からの要求を認識した プロセッサ5200、6200は、共通メモリ領域をキャッシュ3400への1重要さに変更する。次に、プロ セッサ3200は、プロセッサ4200の処理を引き数 ぐ為に、プロセッサ4200用額域の制御権をプロセッ サ3200に切り着える(ステップ953)。 これらの 処理により、制御権の切り替えが完了し、プロセッサ3

(0048) 図8は、監察が発生したコントローラ40 00の復旧処理を示すフローチャートである。コントローラ4000の際書部位が交換(ステップ971)され ると、プロセッサ4200は、共通メモリ領域3410 20 を用いて全プロセッサに提旧開動を伝達する (ステップ

200は通常の1、O処理を再開する (ステップ95

972). プロセッサ3200, 5200, 6200 この復旧開始の伝達を受けて(ステップ955)。 それぞれのコントローラのデータ転送制御館にキャッシ ュ3400と4400への2歳費きを指示すると共に、 共高メモリ領域3410、4410を用いて、処理終了

の応答をプロセッサ4200に通信する(ステップ95 6)。この終了報告を全プロセッサから受知(ステップ 973)したプロセッサ4200は、キャッシュ440 0のデータ同復を行う(ステップ974)。データ回復 が完了すると、共識メモリ領域3410、4410を用いて、プロセッサ3200に復旧完了を伝達する(ステ

77975). 【9.0.4.9】 関7において、この完了適知を受けた(ス トップ958)プロセッサ3200は、プロセッサ42

0 の月前域の初間接をプロセッサイ2 0 0 に似旧(ステップ9 5 9)とせ、共通メモリ間域を用いて、制御権の 質問をプロセッサイ2 0 0 に伝達する (ステップ9 6 0) 、M.8 において、この伝達を受けた(ステップ97 プロセッサ4200は、1.10処理を解除させる (ステップタ77)。

【0050】商、以上の実施側においては、コントロー う毎にプロセッサ、ホストー、下制製器を1つ行った例 を示したが、これらの数は任意でも、ホストコンピュー タからのコマンドを受け取ったプロセッサが、担当プロ セッサに処理要求を伝達することにより、同様に実現で

100511 #t #embanonsets Vote である。 特に、特定プロセッサをホットスタンパイで をである。 特に、特定プロセッサをホットスタンパイで - 50 動作させる場合には、キャッミュ和城をホットスタンパ

2003 04 24 10 53

記憶サブシステム

特別平9~146842

イのプロセッサには割り当てないことにより、キャッシ ュモ有効に利用することができる。又、プロセッサの負 毎に応じてダイナミックに変更することも可能である。 ユーザの指定により分割を行うか、プロセッサの負荷に 応じて変更を行うかの指示は、本実施例では、 マンドにより行うが、パネルといった装置を接続し、そ

マンドにより日 3m. ハドルというに表面を提成し、で こから人力する形を取っても、むろん良い。 {0052} つぎに、コントローラのキャッシュの動的 割当の実現方式について、以下、説明する。まず、キャ ッシュの管理方式について、図9を用いて規則する。 【0053】 ブロセッサ毎に持つデータ格納エリアは、 セグメント983と呼ばれる管理単位に分割されてい る。セグメントは、セグメント毎にセグメント管理プロック981(以下SGCBという。)をデータ管理情報 内に持ち、セグメントを管理する情報とセグメントアド レスが格納されている。又、これらのSGCBは、その セグメントの属性によって、ダーディキュー980とク

リーンキュー982という2つキューに分けられて接続 されている。ダーティキュー980には、ディスク米反 映のライトデータを格納しているセグメントのSGCB が接続されており、それ以外のSGCBは、クリーンキ -982に接続されている。 【0054】キャッシュの動的割当を実現するために、 「ロセッサ谷の食荷体報を共通メモリ領域に持つ。この 食荷体報として、例えば、キャッシュ内のクリーンSG CB量を用いる。各プロセッサは、SGCBのクリー

ン、ダーティ間のキュー遷移契機に、この情報を更新す る。プロセッサは、例えば、1分といった一定周期でこ の情報を参照にいき、キャッシュを共有しているプロセッサ内で最も負荷の低いプロセッサのクリーンキューか ら最も負荷の高いプロセッサのクリーンキューへ、その **食器が同じになるまでSGCRと管理セグェントをは行** させる。この際、使用中のSGCBは、移行対象外とす る。移行の際は、SGCBの格納データ情報はクリアす る。この移行の間は、プロセッサ通信を用いて、移行を 行うプロセッサの 1、「O 処理はどめる。

【0055】また、以上の実施側においては、2分のコ ントローラ間でキャッシュを共有し、各々、 3コントロ ーラのキャッシュに 2 重書きする例を示したが、キャッ シュ価値がプロセッサ毎に分裂されていれば、そのキャ ッシュの共有化方式、多吸着さが式は、任意の方式でも、同様に実現できる。

【0056】キャッシュ多数書きの例を図10に示す。 (1) は、装置全体でキャッシュを共有しあい、2重値 さずる方式である。つまり、プロセッサ3200はキャ マラの水である。 7490 を用いて、プロセッサイ20 のはキャッシュ4400、5400を用いて、プロセッサイ20 りはキャッシュ4400、5400を用いて、プロセッサ5200はキャッシュ5400、6400を用いて、 プロセッサら200はキャッシュ6400、3400を 用いて2重要をを行っている。

【0057】(2)は、協議全体でキャッシュを共有し あい、全キャッシュに多重書きする方式である。つま り、プロセッサ3200、4200、5200、620 Dは、それぞれキャッシュ3400、4400、540 の、6400を用いて、多数書きを行っている。このケ

一スにおいて、コントローラが障害となった場合は、キ ャッシュを共有しているプロセッサ間でもっとも負荷の 低いプロセッサが、微害コントローラ担当論理ポリュー 低いプロモックが、陶香コントローラ和当島東走りュー のの気度を引き起く。これらかヤースにおいては、任皇 のプロセッナが開着コントローラ相当島東ボリュームの 見理を引き継げるように、ディスク側のデーツバスを 装置内の全ディスク装置、モフントローラで下着のバス に接続しておく、もちろん。これらの多葉巻を方式を拡

電内で設在させることも可能である。これらの多数書き 方式の指定は、共通メモリ能域3410、4410に多 重書き情報を持ち、各々のプロセッサ3200、420 0、5200、6200が、この情報を元に、書き込み データの転送方式をデータ転送制御部3300 0、5300、6300に指示することにより実現でき

20 δ.

【兒明の効果】本兒明によれば、コントローラ及びキャ ッシュメモリを2番化した記憶サブシステムにおいて、 各コントローラにキャッシュメモリの「部及び幕壁ボリ ュームを割り当てることによりキャッシュメモリに対す るコントローラ内のプロセッサ間の排他制御が無くなる ため、複数プロセッサ化による応答性能劣化を妨ぐこと

【0059】また、複数のキャッシュへ名を書きするこ とにより、キャッシュ教育時には、多重書きしている他 キャッシュからディスクに書き込むことができるため、 データロストを防ぐことができる。さらに、コントロー ラ駒吉時にキャッシュメモリの制御を正常なコントロー うに切り替える手段とコントローラ確実から提出する毛 段を設けることにより、システムを無停止で適用するこ

とができる. Date of the second second

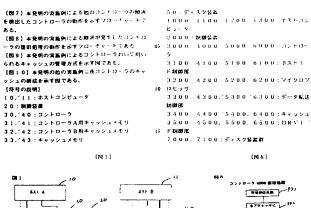
【凶!】 本発明の概要を表す講成はである 【図2】 4発明の実施例である新藤装置の構成料であ

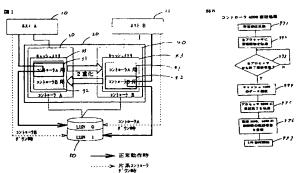
【阿3】本発明の実施例であるコントローラのキャッシ

1の構成を示す物である。 【関4】 本発明の実施例によるコントローラのホストか

らの1 〇姓間の動作を示すフローチャートである。 【図5】 本発明の実施例によるコントローラのキャッシ っ内のデータをディスク装置に集前する動作を示す 一チャートである。

【陽6】本発明の実施例による他のコントローラからめ 理長坂を受けとったコントローラの制御装置の動作を4 50 すフローチャートである。





(M 2) -(V-12) =();·...) =();.!/) 2···· 1210-7 . / 4000 1210-7 . ();...) ?~~ 34t 33t* 1-1 -()-11)·?*** 1200 41) 37()-7 () .U) 7/40 7/00 /300 7/10 121 \$A} 375°5• (), ...) YA

2003 04 24 10 53

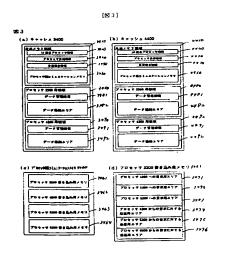
2003 04 24 10 53

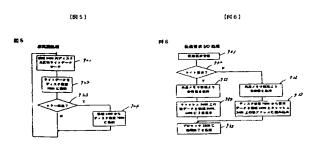
記憶サブシステム

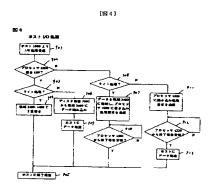
特開平9-146842

記憶サブシステム

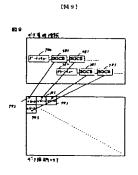
特闘平9-146842



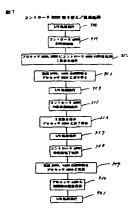




- 10 -

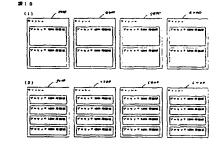


[图7]



. 11 -

[14 | 0]



2003 04 24 10.53

特開平9-146842

- 14 -

2003 04 24 10.53